DOCUMENT RESUME

ED 053 177 TM 000 706

AUTHOR	Pennell, Roger
TITLE	Additive Representations for Two-Dimensional Tables.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-70-29
PUB DATE	Apr 70
NOTE	17p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Algorithms, *Analysis of Variance, *Computer Programs, Goodness of Fit, *Interaction, *Mathematical Models, Mathematics, Multiple Regression Analysis, Statistical Analysis, *Transformations (Mathematics)
IDENTIFIERS	Conjoint Measurement

ABSTRACT

A model and a computer program for performing conjoint measurement is developed. (AG)

ED053177

RESEARCH BULLETIN

RB-70-29

ADDITIVE REPRESENTATIONS FOR TWO-DIMENSIONAL TABLES

Roger Pennell

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
April 1970

TM 000 206

ADDITIVE REPRESENTATIONS FOR TWO-DIMENSIONAL TABLES

Roger Pennell
Educational Testing Service

Abstract

Given that we collect observations A_iP_j from two perfectly crossed factors we may be interested in fitting a model such as $f(A_iP_j) = \alpha_i + \beta_j$. An iterative method for computing the scale values α_i and β_j and the function f is developed. The procedure is relevant to problems of finding monotonic transformations eliminating interaction effects preceding analysis of variance and to the classical conjoint measurement model.

ADDITIVE REPRESENTATIONS FOR TWO-DIMENSIONAL TABLES

Roger Pennell
Educational Testing Service

I. INTRODUCTION

When conjoint measurement was finally exposited in its complete or nearly complete form by Luce and Tukey in 1964, it was heralded by many as the panacea for measurement problems in the behavioral sciences. If not a panacea, certainly the forerunner of an axiomatized, complete measurement system. Indeed certain refinements, simplifications and generalizations followed (Scott, 1964; Krantz, 1964; Roskies, 1965), but the utilization of CM as a tool for research in the behavioral sciences didn't. The reasons for this lack of enthusiasm by the researcher in the field are complex, but probably include the following: (1) only the most sophisticated of readers could wade through the myriad of involved axioms and theorems; (2) the how to do it part of the model was by andlarge lacking. It is the second difficulty to which this paper hopes to make a modest contribution.

CM, as expounded by Luce and Tukey, assumes two sets of "events" or factors, A and P and a weak ordering[1] ($\preceq$) on $A \times P$, where $A \times P$ is the cartesian product of A and P. This is most easily understood in an analysis of variance context where we have two perfectly crossed factors A_i and P_j where $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$, $m,n \geq 2$, and we make observations, or record responses, on each of the cells in the design A_iP_j. We can then generate a weak ordering by lining up the cells,

the A_iP_j, in a nonincreasing sequence. Given this, and certain other conditions discussed below, the axioms assert that there exists three functions ϕ, θ, and Λ such that

$$\phi(A \times P) = \theta(A) + \Lambda(P) \quad . \tag{1}$$

That is, there exist functions which will transform the observations, or responses, and the sets of events, A and P, such that we generate an additive model. In the analysis of variance context, we have found a transformation which permits us to disregard the interaction term and speak of a completely additive model of main effects.

Besides the axiom concerning a weak ordering certain other conditions must be met for (1) to be obtainable: (a) the levels of the factors must be sufficiently finely graded such that for any A_iP_j and A_k for $i \neq k$ there exists a P_ℓ such that $A_iP_j = A_kP_\ell$, (b) given a set of observations $A \times P$ it is necessary that the data conform to certain, essentially transitivity requirements. Formally, given A_i, A_j, A_k and P_ℓ, P_m, P_n, $A_iP_m \geq A_jP_n$ and $A_jP_\ell \geq A_kP_m$ implies $A_iP_\ell \geq A_kP_n$, and (c) an archimedean axiom[2] for ordered sets.

In practice the weak ordering conditions are easy to verify on an empirical set of data; however, the additional conditions enumerated above are not routinely verifiable. Even if we are able to apply the axioms to a finite set of data and deduce that the data do not conform to the axioms, we cannot conclude that an additive representation does not exist, since the axioms of CM are merely sufficient, not necessary. Zinnes (1969) points out another difficulty: we, quite naturally, can't expect the axioms to hold exactly, thus how close do they have to be before we accept them as

being satisfied? There is as yet no statistical theory for determining the goodness of fit of the data to the axioms.

This paper will outline a straightforward approach to achieving CM largely in the context of analysis of variance. The approach may fail for some particular data set in hand; however, from the results decisions can be made as to the utility of the additive representation obtained.

In analysis of variance transformations of the data are almost never for reasons other than achieving homogeneity of variance. Finding a transformation that will yield some model as a strictly additive function of the main effects and an error term, thus eliminating an interaction effect, is rarely considered, even though the existence of such an interaction effect may simply be a result of the scale upon which the factors were measured. In any case a transformation is considered admissible only if it is monotonic, thus preserving the ordering of the cell means. We shall deal with the most general class of monotone functions for achieving an additive representation.

Shepard (1962) was the first to explicitly state the notion of nonmetric monotonicity as a criterion for admissible transformations of observed data. That is, a transformation is considered admissible only if the ordinal properties of the original data are maintained after transformation. Certainly in the context of analysis of variance, it is inconceivable to consider transformations which may invert the order of the cell means (Winer, 1962). Kruskal (1964a, 1964b) implemented Shepard's original notions into a powerful algorithm for resolving a set of data into its dimensional components. The crux of Kruskal's program is the generation of a monotone regression of reproduced data on original data. At each step in the program the data points, in a space of given dimensionality, are altered slightly

to maximize this regression. Since only the ordinal characteristics of the original data are of interest (more appropriately, the ordinal relations are the maximal information obtainable), we can't strictly talk of maximizing a regression (that is, doing arithmetic on ordinal numbers). Thus, Kruskal essentially discards the original data while preserving only its rank ordering. Symbolically given a sequence of data $D = \{d_i\}_{i=1}^{N}$, let σ be a permutation of the first N integers such that $d_{\sigma(1)} \geq d_{\sigma(2)} \geq \cdots \geq d_{\sigma(N)}$. The function, σ, is the only characteristic of the data retained. The monotone regression problem is then solved by applying the same function to the reproduced data and seeking a partition which renders it also nonincreasing. For details of this procedure see Kruskal (1964b).

II. METHOD

We shall use essentially the same approach in finding a transformation, $f(A_iP_j)$, which is exactly monotonically related to the original A_iP_j. Explicitly we shall try to fit a model of the form

$$f(A_iP_j) = \alpha_i + \beta_j + \varepsilon_{ij} \quad . \tag{2}$$

To do this we shall use a measure similar to Kruskal's stress, S :

$$S = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \varepsilon_{ij}^2 \Big/ \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j)^2 \right]^{1/2} \quad . \tag{3}$$

Minimizing (3) is equivalent to finding a set of α_i and β_j which reproduce, in an additive fashion, a transformation of the original data. We shall assume, without loss of generality, that $\sum_{j=1}^{m} \beta_j = 0$. Taking the

partial derivative of (3) with respect to the α_i and β_j, and noting that minimizing S is equivalent to minimizing S^2 we obtain

$$(4a) \qquad \frac{\partial S^2}{\partial \alpha_v} = mK\alpha_v - K \sum_{j=1}^{m} f(A_i P_j) - mL\alpha_v = 0$$

$$(4b) \qquad \frac{\partial S^2}{\partial \beta_\ell} = (L = K)[\sum_{i=1}^{n} \alpha_i + n\beta_\ell] - L \sum_{i=1}^{n} f(A_i P_j) = 0 \quad ,$$

where $K = \sum_{i=1}^{n} \sum_{j=1}^{m} \epsilon_{ij}^2$ and $L = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j)^2$. Since equation (4) is a rather complicated quadratic we choose to use a gradient method (Kunz, 1957) in order to minimize $\partial S^2/\partial \alpha_v$ and $\partial S^2/\partial \beta_\ell$. That is to say, given a set of initial approximations to start the process, say α_{i0} and β_{j0}, the estimates of the final α_i and β_j after the k-th iteration will be

$$(5) \qquad \alpha_{ik} = \alpha_{ik-1} - \lambda \frac{\partial S^2}{\partial \alpha_{ik-1}}$$

$$\beta_{jk} = \beta_{jk-1} - \lambda \frac{\partial S^2}{\partial \alpha_{ik-1}} \quad .$$

As an aside we can note that heuristically one can picture (5) as hunting in a space of nm dimensions (the parameter space of α_i s and β_i s) for a point providing a minimum for S. We can consider the point having coordinates α_{i0}, $i = 1,2,\ldots,n$, β_{j0}, $j = 1,2,\ldots,m$ as lying on a hypersurface of constant S_0. By evaluating (5) we move in a direction perpendicular to the hypersurface, inward towards the point S. We move inward until we just graze another hypersurface of constant, say S_1, reevaluate (4) and again move inward toward S. Ultimately the process should converge to a minimum for S.

We need to estimate two quantities in order to proceed. First, it would be helpful to have a reasonable first approximation to the final solution for the α_i and β_j. We should be able to obtain such an approximation by assuming the (A_iP_j) are additive and letting α_{i0} and β_{j0} correspond to the least squares estimates

$$\alpha_{i0} = \frac{1}{n} \sum_{j=1}^{n} (A_iP_j) - C \tag{6a}$$

$$\beta_{j0} = \frac{1}{m} \sum_{i=1}^{m} (A_iP_j) \quad , \tag{6b}$$

where C represents the average α effect.

The second quantity we need is λ, commonly known as the step size. In our case it is rather simple to compute if we observe that for any iteration, $k + 1$, we are trying to find a λ which makes

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \varepsilon_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_{ik} - \lambda \frac{\partial S^2}{\partial \alpha_{ik}} + \beta_j - \lambda \frac{\partial S^2}{\beta_{jk}} - f(A_iP_j))^2 \tag{7}$$

$$= KS = T$$

as small as possible. Thus, the needed value of λ is given by $dT/d\lambda = 0$, or, dropping the iteration subscripts and letting $V_i = \partial S^2/\partial \alpha_i$ and $Q_j = \partial S^2/\partial \beta_j$,

$$\lambda = \tag{8}$$

$$\frac{m \sum_{i=1}^{n} \alpha_i V_i + \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i Q_j + n \sum_{j=1}^{m} \beta_j Q_j + \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j V_i - \sum_{i=1}^{n} \sum_{j=1}^{m} f(A_iP_j) V_i - \sum_{i=1}^{n} \sum_{j=1}^{m} f(A_iP_j) Q_j}{m \sum_{i=1}^{n} V_i^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{m} Q_j V_i + n \sum_{j=1}^{m} Q_j^2} \quad .$$

In summary, given a set of data to be measured conjointly we (1) string the data out in a vector (for instance, take each column, one at a time, and string it into a single vector) and find the function σ , (2) find the initial approximations α_{i0} and β_{j0} , (3) string the numbers $\alpha_{i0} + \beta_{j0}$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$, into similarly arranged vector as in step 1 and apply the function σ , (4) find the function f , (5) solve equations (3), (4) and (8) and obtain improved estimates of α_i and β_j and repeat starting from step 3.

III. EXAMPLES

In order to illustrate the above outlined algorithm two examples will be presented. The first is some data taken from Winer (1962, p. 245). These data are supposed to represent a 2 x 2 analysis of variance with two observations per cell. The sums of the observations are presented in Table 1.

- - - - - - - - - - - - - -
Insert Table 1 about here
- - - - - - - - - - - - - -

The F -test for the interaction effect is significant at better than $p = .01$ and the associated sum of squares is 950.56. Winer conjectured that a square-root transformation would remove the sum of squares (SS) due to interaction, and, in fact, such a transformation reduced it to .30. Table 2 presents the same data after they have been transformed by a computer program designed to carry out CM .

- - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - -

In approximately one second of 360/65 CPU time the program was terminated after 10 iterations and $S = .7 \times 10^{-6}$. Sums of squares for interaction was less than $.49 \times 10^{-7}$. Values of S_i as well as the final scale values are also presented. We should expect $S = 0.0$ for data in which none of the profiles of plotted cell means cross between effects. These data were probably concocted for illustrative purposes by Winer in order to show that there often exist transformations on the original scale of measurement which render an essentially additive model. The computer program, however, made no assumptions about the form of the transformation except that it be monotonic, and recovered an essentially square-root transformation (linearly transformed) such that the plot of the cell means looks almost exactly like Winer's (p. 247) except that the SS for interaction is zero to five decimal places.

The second example represents some data collected by Leibowitz and Bourne (1956) attempting to explore the conditions under which either retinal image or shape constancy obtain. They varied the degree of luminance and the duration of exposure obtaining the data presented in Table 3.

- - - - - - - - - - - - - - -
Insert Table 3 about here
- - - - - - - - - - - - - - -

The data indicate that as luminance or exposure is increased shape constancy tends to obtain, and conversely under minimal viewing conditions (near threshold) retinal image tends to dominate.

The data were input to the CM Program and produced the results in Table 4. The solution was obtained after an arbitrary 20 iterations with

- - - - - - - - - - - - - -
Insert Table 4 about here
- - - - - - - - - - - - - -

$S = .311 \times 10^{-2}$ and sum of squares for interaction equal to $.719 \times 10^{-4}$.

If we force the data to admit an additive structure it can be seen that the experimental design is highly redundant in that luminances of .1 and 1.0 and exposures of .01 and .05, and .5 and .75 produce highly similar perceptions. Further, we have found a monotone transformation of the values expressing duration of exposure and amount of luminance which yields a very close additive model.

IV. DISCUSSION

After completing the development of the model and producing a computer program to perform CM it came to the author's attention that a very similar, although slightly more sophisticated, approach had been devised in a book by Roskam (1968). His approach proceeds by a direct minimization of an equation similar to (3) and appears to produce results similar to those reported here except for those cases extremely degenerate in form. Young (1969) has also reported an algorithm for doing polynomial CM in N -space which is a generalization (using a different algorithm) of the results reported here.

The generalization of our approach to N -dimensional scale values is straightforward, though certainly coupled with some risk. In practice we surely would not let the total number of scale values exceed mn , and certainly should explore the tenability of a one-dimensional fit first.

References

Krantz, D. H. Conjoint measurement: The Luce-Tukey axiomatization and some extensions. Journal of Mathematical Psychology, 1964, 1, 248-277.

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129. (a)

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27. (b)

Kunz, K. S. Numerical analysis. New York: McGraw-Hill, 1957.

Leibowitz, H., & Bourne, L. E., Jr. Time and intensity as determiners of perceived shape. Journal of Experimental Psychology, 1956, 51, 277-281.

Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

Roskam, E. E. Ch. I. Metric analysis of ordinal data in psychology. Netherlands: Vain-Voorschoten, 1968.

Roskies, R. A measurement axiomatization for an essentially multiplicative representation of two factors. Journal of Mathematical Psychology, 1965, 2, 266-276.

Scott, D. Measurement structures and linear inequalities. Journal of Mathematical Psychology, 1964, 1, 233-247.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika, 1962, 27, 125-139, 219-246.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Young, F. W. Polynomial conjoint analysis of similarities: Operations of a specific algorithm. L. L. Thurstone Psychometric Lab. Report 77, 1969.

Zinnes, J. L. Scaling. Annual Review of Psychology, 1969, 20, 447-478.

Footnotes

[1]Formally, for $\leq$ to be a weak ordering the following must obtain for q, r and $s \in A \times P$: (1) $q \geq q$ holds for all q, (2) $q \geq r$ and $r \geq s \Rightarrow q \geq s$, (3) either $q \geq r$ or $r \geq q$ or both.

[2]Briefly, an archimedean axiom generally requires that for arbitrary A_iP_j and A_kP_ℓ there exists an integer n such that $nA_iP_j \geq A_kP_\ell$.

Table 1

Sum of Two Observations in a Two-way
Analysis of Variance
(from Winer, 1962, p. 245)

	b_1	b_2	b_3
a_1	1.0	26.0	47.0
a_2	18.0	62.0	95.0
a_3	64.0	134.0	196.0

$SS_{AxB} = 950.56$

Table 2

Data from Table 1 after Conjoint Measurement, Scale Values, and Values of S During Iteration

		b_1	b_2	b_3
	Scale Values	27.658	75.251	111.424
a_1	-48.037	-20.378	27.214	63.387
a_2	-11.863	15.795	63.387	99.561
a_3	59.900	87.559	135.151	171.324

Iteration Number	1	2	3	4	5	6	7	8	9	10
S	$.129\times10^{-1}$		$.144\times10^{-2}$		$.16\times10^{-3}$		$.178\times10^{-4}$		$.2\times10^{-5}$	
		$.432\times10^{-2}$		$.481\times10^{-3}$		$.533\times10^{-4}$		$.59\times10^{-5}$		$.7\times10^{-6}$

Table 3

Mean Ratios of Major Axis Length to Minor Axis Length of Elipses Matched to a Standard (.5) at Varying Levels of Duration and Luminance (Leibowitz and Bourne (1956), p. 278)

Exposure (sec.)	Luminance (Millilamberts)		
	.01	.1	1.0
.01	.486	.524	.515
.05	.503	.528	.517
.10	.522	.566	.570
.25	.544	.608	.692
.50	.570	.688	.802
.75	.575	.670	.790
1.00	.590	.737	.842

Table 4

Data from Table 3 after
Conjoint Measurement

Scale Values	.556	.636	.642
-.113	.443	.526	.526
-.113	.443	.526	.526
-.030	.526	.607	.613
.052	.607	.688	.698
.064	.620	.698	.708
.064	.620	.698	.708
.075	.631	.708	.718